

# ELASTIC IMAGE REGISTRATION WITH EXACT MASS PRESERVATION

JAROSŁAW WLAZŁO, ROBERT FESSLER, RENÉ PINNAU, NORBERT SIEDOW, AND OLIVER TSE

**ABSTRACT.** We establish a new framework for image registration, which is based on linear elasticity and optimal mass transportation theory. We combine these two arguments in order to obtain a PDE constrained optimization problem that is analytically investigated and further discretized with the finite difference method and solved by an inexact SQP algorithm. This requires to solve in each step a large sparse linear system, which has a saddle point form. Motivated by stability arguments we use a fully staggered grid for the discretization of the displacement vector field. Artificial and real world examples are presented to underline the numerical robustness of the method.

**Keywords.** Image registration; optimal transportation; mass preservation; linear elasticity.

**AMS Classification.** 92C55, 49J20, 49K20, 65M55, 65K10

## 1. INTRODUCTION

Image registration is one of the most challenging problems in image processing. Especially in medicine, which is the main direction of our interest, this process has an extremely high significance. It allows for the incorporation of different image information to facilitate and improve surgery planning or therapy procedure. Recently, as a consequence of the advanced development of imaging techniques it has become possible to perform a minimal invasive operation using real live imaging. However, in order to make the image information usable for surgeons image registration becomes essential. Moreover, an important issue is also the quality of the registration process since its results may dictate the success of a surgery.

The image registration problem may be expressed as follows. Let  $R$  and  $T$  be any two given images represented by their intensity functions on a domain  $\Omega \subset \mathbb{R}^d$ , where  $d = 2$  corresponds to plane images and  $d = 3$  to 3D images. In the following we will call  $R$  the *reference* and  $T$  the *template* or *moving* image. We wish to find a common geometric reference frame between these two images, i.e., the task is to find a transformation  $\phi : \Omega \rightarrow \Omega$  such that the transformed template image  $T_\phi := T \circ \phi$ , resembles the reference image as much as possible. Naturally, this resembling can be understood in a range of different ways, depending on what kind of similarity measure we take into account. A precise mathematical formulation and physical reasoning of our choice is described in next section.

One can observe an increasing number of publications in the field of image registration over the past decades. Various models have been established and many numerical methods have been implemented [33]. The presented approaches differ from rigid to non-rigid [9, 45], landmarks based to intensity based [40, 18], or parametric to non-parametric ones [30, 36]. The method of choice strongly depends on the specific properties of the problem. It is rather impossible to indicate an exceptional procedure which succeeds in every case. Each of them has its own advantages and drawbacks.

A most common treatment of the task at hand is based on the following variational formulation. Find a transformation  $\phi$  minimizing the total energy

$$(1) \quad \mathcal{J}_\alpha[\phi] := \mathcal{D}(R, T_\phi) + \alpha \mathcal{S}(\phi - \phi_{\text{ref}}),$$

where  $\mathcal{D}$  corresponds to a dissimilarity measure quantifying the difference between the transformed template,  $T_\phi$ , and the reference,  $R$ . In this work we are mostly interested in mass preserving transformations [41]. In this case  $\mathcal{D}$  takes the form

$$(2) \quad \mathcal{D}(R, T_\phi) := \int_{\Omega} (T_\phi(x) \det(\nabla \phi(x)) - R(x))^2 dx.$$

In (1),  $\mathcal{S}$  is a regularizing term which eliminates unwanted transformations,  $\phi_{\text{ref}}$  is a favoured transformation (typically set to identity) and  $\alpha > 0$  is a regularization parameter that controls the influence of  $\mathcal{S}$ .

In [34] (cf. [44]) the authors showed that image registration based solely on  $\mathcal{D}$  is an ill-posed problem. One can easily understand this observation by realizing that image registration is a process that tries to determine a vector field describing the transformation, while the available data is a scalar field. This is the reason for introducing  $\mathcal{S}$  as a Tikhonov regularizing term that ‘convexifies’ the functional in (1). Well-posedness of the regularized problem (1) depends naturally on the choice of  $\mathcal{D}$  and  $\mathcal{S}$ . In most of the standard settings, one can show existence of optimal solution, see [42]. On the other hand, showing uniqueness is rather difficult or even impossible.

Among possible choices of dissimilarity measures, there are the sum of squared differences (SSD) [26], mutual information [32], cross correlation [31] and others. Regarding regularizers, one typically uses diffusive [16, 45], curvature [17, 18] or elastic energies [24] (see also [33] and the references therein). If the imaged objects are human tissues, the elastic regularizer becomes reasonable and at times even essential. For small deformations the linear elastic model has been successfully utilized in image registration. In order to handle large deformations several approaches have been proposed. The simplest one makes use of pre-processing with a large scale affine transformation [2]. One may also use nonlinear or hyperelasticity theory [21, 5, 41]. A completely different strategy, proposed by Christensen et. al. [8], uses arguments from the theory of fluid dynamics. In this formulation, the transformation is described by the velocity field rather than the displacement. Other additional constraints may also be introduced to guarantee a desired property of the transformation, e.g., volume preservation [23] or local rigidity [35].

A different but related strategy to the one described in this paper was presented by Zhu et. al. [48], and further investigated by Rehman in [39]. These works are based on  $L^2$  optimal mass transportation [15], where the similarity between the two given images is measured locally by a partial differential equation describing the mass preservation property. Thus the image registration problem is expressed by the constrained optimization problem

$$(3) \quad \begin{aligned} \min \quad & W(\phi) := \int_{\Omega} |\phi(x) - x|^2 \rho_R(x) dx \\ \text{subject to} \quad & \det(\nabla \phi(x)) \rho_T(\phi(x)) = \rho_R(x) \quad \text{for } x \in \Omega, \\ & \phi(x) = x \quad \text{for } x \in \partial\Omega, \end{aligned}$$

where  $\rho_R$  and  $\rho_T$  represent the mass densities for  $R$  and  $T$  respectively, and are derived from the input images. This eliminates the need of a regularization term and makes the image registration problem completely parameter-free. Our approach is motivated by similar arguments, since we believe that the input data are true representations of the mass densities  $\rho_R$  and  $\rho_T$  and there is no additional mass flow in the time interval when the shots  $R$  and  $T$  are made. However, here, instead of the classical  $L^2$  transportation cost we minimize the linear elastic strain energy [10], which in terms of displacement  $u$  has the following form

$$(4) \quad \mathcal{S}(u) := \int_{\Omega} \frac{\mu}{4} \sum_{i,j=1}^d \left( \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right)^2 + \frac{\lambda}{2} \left( \sum_{i=1}^d \frac{\partial u_i}{\partial x_i} \right)^2 dx,$$

where  $\mu$  and  $\lambda$  are two positive Lamé parameters describing the mechanical properties of the imaged object. Note, that the objective functional in (3) does not take into account derivatives of the displacements. As a consequence, the minimizing direction may lead to irregular solution. In contrast, the elastic potential energy rules out non-regular mappings, thereby providing a more smooth result.

The paper is organized as follows. In Section 2 we describe the mathematical set-up and prove an existence result for the problem. Section 3 is devoted to describing the discretization process and the assembling of the discrete analogous to the continuous image registration problem. We further provide a detailed description of the discretization in Appendix A. In Section 4 we present the optimization strategy and formulate an algorithm to solve the image registration problem. In Section 5 we present two numerical examples and finally conclude in Section 6.

## 2. MATHEMATICAL FORMULATION AND EXISTENCE PROOF

We now describe a parameter-free approach to the image registration problem in 2D. Let  $\Omega \subset \mathbb{R}^2$  be a sufficiently smooth domain and  $\mathcal{M}$  denote a suitable function space of deformations over  $\Omega$ . We will specify the particular choice of  $\mathcal{M}$  later on, but we always require  $\mathcal{M}$  to be a closed subset of a reflexive Banach space  $X$ . Notice that due to the linear elasticity assumption we have the following relation between the total deformation  $\phi$  and the displacement field  $u$ :  $\phi = \phi_{\text{ref}} + u$ , where  $\phi_{\text{ref}}$  is a reference deformation for which the strain energy vanishes. In our case we naturally set  $\phi_{\text{ref}}$  to be the identity map  $id$ , i.e.,  $\phi_{\text{ref}}(x) = id(x) = x$  for all  $x \in \Omega$ . For any given density functions  $\rho_R, \rho_T : \Omega \rightarrow \mathbb{R}$  having equal mass, i.e.,

$$(5) \quad \int_{\Omega} \rho_R(x) dx = \int_{\Omega} \rho_T(x) dx,$$

we define the following constrained optimization problem:

$$(P_{\phi}) \quad \begin{aligned} \min_{\phi \in X} \quad & \tilde{S}(\phi) := \mathcal{S}(\phi - \phi_{\text{ref}}) \\ \text{subject to} \quad & \tilde{c}(\phi; \rho_R, \rho_T) := \det(\nabla \phi) \rho_T \circ \phi - \rho_R = 0 \quad \text{in } \Omega \\ & \phi(x) = x \quad \text{on } \partial\Omega. \end{aligned}$$

In order to simplify the considerations we put here several assumptions.

**Assumption 1.** The domain  $\Omega$  is an open and bounded subset of  $\mathbb{R}^2$  with Lipschitz boundary  $\partial\Omega$ .

**Assumption 2.** The image intensity functions are uniformly positive, i.e., there exists  $\delta > 0$  such that

$$(6) \quad \rho_R(x) \geq \delta, \quad \rho_T(x) \geq \delta \quad \text{for all } x \in \Omega.$$

Additionally we require  $\rho_R$  to be continuous, and  $\rho_T$  to be continuously differentiable, i.e.,

$$(7) \quad \rho_R \in C^0(\bar{\Omega}), \quad \rho_T \in C^1(\Omega) \cap C^0(\bar{\Omega}).$$

With the above assumptions we may reformulate the mass preserving condition in  $(P_{\phi})$  to an equivalent form with a separated determinant part, which reads

$$(8) \quad \bar{c}(\phi; \rho_R, \rho_T) := \det(\nabla \phi) - \frac{\rho_R}{\rho_T \circ \phi}.$$

In order to shorten the notation we will skip the second and third arguments and simply write  $\bar{c}(\phi) = \bar{c}(\phi; \rho_R, \rho_T)$ . It is easy to see that with (6)  $\bar{c}$  is well defined and has the same null space as the mass preserving constraint  $\tilde{c}$  in  $(P_{\phi})$ . Let us define  $\mathcal{M}$  to be a subset of  $X$  that contains all feasible deformations given by

$$(9) \quad \mathcal{M} := \left\{ \phi \in X : \bar{c}(\phi) = 0 \text{ in } \Omega, \phi(x) = x \text{ on } \partial\Omega \right\}.$$

Note, that the double symmetric gradient of the reference deformation,  $\phi_{\text{ref}} = id$ , is the identity matrix  $I$  and thus  $\tilde{S}$  becomes

$$(10) \quad \tilde{S}(\phi) = \frac{1}{2} \int_{\Omega} \frac{\mu}{2} \left( \nabla \phi + \nabla \phi^{\top} - I \right) : \left( \nabla \phi + \nabla \phi^{\top} - I \right) + \lambda \left( \text{tr}(\nabla \phi - I) \right)^2 dx$$

Finally, we formulate the existence theorem for our elastic mass preserving image registration problem  $(P_{\phi})$ . It has been shown in [12] that, under certain additional assumptions, one has the existence of mass preserving mappings in  $C^1(\Omega; \mathbb{R}^d)$ . However, in our case we additionally require the solution space to be complete with respect to a norm derived from  $\mathcal{S}$ . Therefore, we naturally choose  $X$  to be the Sobolev space  $X = H^1(\Omega; \mathbb{R}^2)$ .

**Theorem 1.** *Let  $\Omega \subset \mathbb{R}^2$  satisfy Assumption 1 and  $\rho_R$  and  $\rho_T$  be two real functions on  $\Omega$  satisfying Assumption 2. Then the problem  $(P_{\phi})$  has at least one optimal solution in  $X$ .*

In order to prove the above theorem we proceed according to the direct method of calculus. We split the proof into three parts: 1) weak lower semicontinuity (w.l.s.c.) of  $\tilde{S}$  in  $X$ , 2) coercivity of  $\tilde{S}$  on  $\mathcal{M}$ , 3) weak sequentially closedness of  $\mathcal{M}$  in  $X$ . If the three conditions are satisfied, the statement of Theorem 1 directly follows.

**Lemma 1.** *The functional  $\tilde{\mathcal{S}}$  defined in (10) is weakly lower semicontinuous in  $X$ .*

*Proof.* It is straightforward to observe that the functional  $\tilde{\mathcal{S}}$  is a composition of linear and convex operators and is thus convex. Additionally, it is continuous with respect to the strong topology on  $X$  and in consequence also weakly lower semicontinuous.  $\square$

We proceed with step 2) and show that  $\tilde{\mathcal{S}}$  satisfies

$$\tilde{\mathcal{S}}(\phi) \rightarrow \infty \quad \text{as} \quad \|\phi\|_X \rightarrow \infty.$$

For  $d = 2$  this is a direct consequence of Korn's inequality [14].

**Proposition 1** (Second Korn's inequality). *Let  $\Omega \subset \mathbb{R}^2$  satisfy Assumption 1. For a displacement field  $u: \Omega \rightarrow \mathbb{R}^2$  we define the linear strain tensor*

$$\epsilon_{ij}(u) := \frac{1}{2}(\partial_i u^j + \partial_j u^i), \quad i, j \in \{1, \dots, d\}.$$

*Then, there exists a positive constant  $\beta = \beta(\Omega)$  such that it holds*

$$(11) \quad \sum_{i,j=1}^2 \|\epsilon_{ij}(u)\|_{L^2}^2 \geq \beta \|\nabla u\|_{L^2}^2, \quad \text{for all } u \in X_0 := \{u \in X : u = 0 \text{ on } \partial\Omega\}.$$

**Lemma 2.** *The functional  $\tilde{\mathcal{S}}$  defined in (10) is coercive on  $\mathcal{M} \subset X$ .*

*Proof.* Set  $\mathcal{N} := \{\phi_{\text{ref}} + u : u \in X_0\}$  and for  $\phi \in \mathcal{N}$  define  $u_\phi \in X_0$  with  $\phi = u_\phi + \phi_{\text{ref}}$ . Then, we have

$$\tilde{\mathcal{S}}(\phi) = \int_{\Omega} \mu \sum_{i,j=1}^d \epsilon_{ij}(u_\phi)^2 + \frac{\lambda}{2} \left( \sum_{i=1}^d \partial_i u_i \right)^2 dx \geq \mu \int_{\Omega} \sum_{i,j=1}^d \epsilon_{ij}(u_\phi)^2 dx \geq \mu \beta(\Omega) \|u_\phi\|_X^2,$$

where the last inequality follows from (11) and the Poincaré inequality. Observe that  $\mathcal{M} \subset \mathcal{N}$  and since  $\phi_{\text{ref}}$  is a constant element in  $X$  the claim follows.  $\square$

Next, we show that  $\mathcal{M}$  is weakly sequentially closed in  $X$ . It is well known that for  $p \geq d$  the Jacobian map

$$\det D : W^{1,p}(\Omega; \mathbb{R}^2) \rightarrow L^{p/d}(\Omega); \quad (\det D)(\phi) := \det(\nabla \phi),$$

is continuous if we endow both spaces with the strong topology. For  $p > d$  we even have continuity in the weak topology. In fact, in two dimensions the special case  $p = d$  is sufficient to show existence.

**Remark 1.** Unfortunately when  $d = 3$  the above consideration requires one to set  $p > 3$ . However, since we wish to stay in Hilbert spaces it is necessary to require more regularity from the functional.

In the following we utilize a classical result on the weak continuity of Jacobian determinants [11, 38].

**Proposition 2.** *Let  $\Omega \subset \mathbb{R}^2$  satisfy Assumption 1. Choose a sequence  $\{\phi_k\} \subset X$  such that  $\phi_k \rightharpoonup \phi$  for some  $\phi$  in  $X$ . Then,  $\det(\nabla \phi_k) \rightharpoonup^* \det(\nabla \phi)$  in the sense of measures, i.e.,*

$$\int_{\Omega} \det(\nabla \phi_k) \psi dx \rightarrow \int_{\Omega} \det(\nabla \phi) \psi dx \quad \text{for all } \psi \in \mathcal{C}_0^\infty(\Omega).$$

With Proposition 2 at hand, we can now prove weak sequentially closedness of  $\mathcal{M}$ .

**Lemma 3.** *Let  $\Omega \subset \mathbb{R}^2$  satisfy Assumption 1 and  $\rho_R, \rho_T$  satisfy Assumption 2. Then, the mass preserving manifold  $\mathcal{M}$  is weakly sequentially closed in  $X$ .*

*Proof.* Let  $\{\phi_k\} \subset \mathcal{M}$  be a sequence such that  $\phi_k \rightharpoonup \phi$  for some  $\phi \in X$ . Due to the results in [12],  $\mathcal{M}$  is non-empty and thus such a sequence exists. We want to show that  $\phi \in \mathcal{M}$ , i.e., it holds  $\det(\nabla \phi) = \rho_R/\rho_T \circ \phi$  in  $\Omega$  and  $\phi = \phi_{\text{ref}}$  on  $\partial\Omega$ .

The second condition is trivial, hence let us prove the first one. Due to Sobolev's embedding theorem, we have the strong convergence  $\phi_k \rightarrow \phi$  in  $L^2(\Omega; \mathbb{R}^2)$ . Consequently,  $\{\phi_k\}$  possesses

a a.e. converging subsequence  $\{\phi_{k_i}\} \subset \{\phi_k\}$  such that  $\phi_{k_i}(x) \rightarrow \phi(x)$  for a.e.  $x \in \Omega$ . From Assumption 2 we further obtain

$$\rho_R/\rho_T \circ \phi_{k_i} =: F_i \longrightarrow F := \rho_R/\rho_T \circ \phi \quad \text{for a.e. } x \in \Omega.$$

Since  $\{\phi_{k_i}\} \subset \mathcal{M}$  we additionally have

$$\det(\nabla \phi_{k_i}) = F_i \quad \text{a.e. in } \Omega,$$

and thus  $\det(\nabla \phi_{k_i}) \rightarrow F$  for a.e.  $x \in \Omega$ .

Observe that the functions  $\det(\nabla \phi_{k_i})$  are uniformly bounded. Indeed, due to Assumption 2,

$$|\det(\nabla \phi_{k_i})| = |\rho_R/\rho_T \circ \phi_{k_i}| \leq (1/\delta) \|\rho_R\|_\infty \quad \text{for all } i \in \mathbb{N}.$$

Therefore, we have that

$$\det(\nabla \phi_{k_i}) \rightharpoonup F \quad \text{in } L^p(\Omega; \mathbb{R}^2), \quad 1 \leq p < \infty.$$

In particular, the above holds for  $L^2(\Omega, \mathbb{R}^2)$ , i.e.,

$$\int_\Omega \det(\phi_{k_i}(x)) \psi(x) dx \rightarrow \int_\Omega F(x) \psi(x) dx \quad \text{for all } \psi \in L^2(\Omega; \mathbb{R}^2).$$

Hence, Proposition 2 gives  $F = \det(\nabla \phi)$  which concludes the proof.  $\square$

Finally, we can prove Theorem 1.

*Proof of Theorem 1.* Let  $\{\phi_k\}$  be a minimizing sequence of  $\tilde{\mathcal{S}}$  in  $\mathcal{M}$ , i.e.,

$$\lim_{k \rightarrow \infty} \tilde{\mathcal{S}}(\phi_k) = \inf_{\phi \in \mathcal{M}} \tilde{\mathcal{S}}(\phi).$$

Due to coercivity (Lemma 2), this sequence must be bounded. Therefore, there is a subsequence  $\{\phi_{k_i}\} \subset \{\phi_k\}$  and  $\phi \in X$  such that  $\phi_{k_i} \rightharpoonup \phi$ . Due to weak sequential closedness of  $\mathcal{M}$  we deduce  $\phi \in \mathcal{M}$ . Finally, the weak lower semicontinuity of  $\tilde{\mathcal{S}}$  (Lemma 1) ensures that  $\phi$  is a desired global minimizer.  $\square$

In the following, we consider an equivalent formulation of  $(P_\phi)$  for displacement fields given by

$$\begin{aligned} (P_u) \quad & \min_{u \in X_0} \mathcal{S}(u) \\ & \text{subject to } c(u) := \tilde{c}(\phi_{\text{ref}} + u; \rho_R, \rho_T) = 0 \quad \text{in } \Omega. \end{aligned}$$

**Remark 2.** As mentioned, in the 3D case the Jacobian determinant involved in the constraints creates some difficulties. One can either leave the Hilbert setting and prove existence in  $W^{1,p}$ , with  $p \geq 3$  or stay in Hilbert space but require more regularity. This can be done, for example, by adding a second-order term to the objective functional.

### 3. DISCRETIZATION

Our numerical method attempts to solve the optimization problem  $(P_u)$  by using a staggered finite difference (FD) discretization on a uniform grid, which is a common practice in computational fluid dynamics (CFD) [13]. The choice of regular grid is in some sense determined by the input image data which, in fact, corresponds to a regular grid of pixel intensities. Moreover, the use of a regular lattice highly reduces the required storage. Since image registration is often applied to very high resolution images, the above argument can be important. Additionally, in general FD methods lead to sparser stencils than their finite element counterpart. Finally, it is well known in CFD that non-staggered discretization schemes may lead to grid-scale oscillations. Since our discrete problem resembles the stationary Navier–Stokes equations, we discretize  $(P_u)$  on a staggered grid in order to stabilize the numerical method and prevent undesired oscillations. The complete and precise discretization process is described in Appendix A.

In this numerical section, we choose the discretize-then-optimize strategy [22], i.e., we first discretize the continuous constrained optimization problem  $(P_u)$  and then derive the required discrete optimality conditions. The solution procedure for the resulting nonlinear system is described in the next section.

Here, we simply clarify that  $\Omega_c$  and  $\Omega_u$  are discrete computational domains of cell centered and edge centered grid points. By  $H_c$  and  $H_u$  we denote the discrete function spaces defined on  $\Omega_c$  and

$\Omega_u$ , respectively (cf. Appendix A.1). Moreover, let  $c^h$  and  $S^h$  denote the constraint and objective functional that act on functions in  $H_u$  and yield values in  $H_c$  and  $\mathbb{R}$ , respectively (cf. Appendix A.3 and A.4). With these notations, we obtain a finite dimensional equality constrained optimization problem with  $n = |\Omega_u|$  unknowns and  $m = |\Omega_c|$  constraints:

$$(P_h) \quad \begin{aligned} & \min_{u^h \in H_u} \quad S^h(u^h) \\ & \text{subject to} \quad c^h(u^h) = 0, \end{aligned}$$

where the boundary conditions are already included in  $S^h$  and  $c^h$ . One can easily observe that  $S^h$  and  $c^h$  are both polynomials in  $u^h$  and thus continuously differentiable. Letting the variable  $p^h \in H_c$  denote the Lagrange multiplier, we construct the Lagrangian function as

$$(12) \quad L(u^h, p^h) = S^h(u^h) + \langle p^h, c^h(u^h) \rangle_{H_c},$$

where  $\langle \cdot, \cdot \rangle_{H_c}$  denotes the inner product in the Hilbert space  $H_c$ . The classical Karush–Kuhn–Tucker (KKT) conditions for optimality then read

$$(13) \quad \begin{aligned} d_u L(u^h, p^h)[v] &:= \delta S^h(u^h)[v] + \langle p^h, \delta c^h(u^h)[v] \rangle_{H_c} & \text{for all } v \in H_u, \\ d_p L(u^h, p^h)[q] &:= \langle q, c^h(u^h) \rangle_{H_c} & \text{for all } q \in H_c, \end{aligned}$$

where the derivatives  $\delta S^h(u^h) \in H_u^*$  and  $\delta c^h(u^h) \in \mathcal{L}(H_u, H_c)$  are defined in (35) and (32), respectively. Using the Riesz representation theorem, the above system reduces to

$$(OC_h) \quad \begin{aligned} A^h u^h + \mathbf{B}^{h,\top}(u^h) p^h &= 0, \\ c^h(u^h) &= 0, \end{aligned}$$

with  $A^h$  defined in (33) and  $\mathbf{B}^h(u^h)$  being the  $m \times n$  Jacobian matrix of  $c^h$  evaluated at  $u^h$ .

**Remark 3.** One easily notices that the divergence operator  $\nabla \cdot$  is a linearization of the Jacobian determinant  $\det D$ . Additionally,  $A^h$  stands for the elliptic Navier–Lamé operator and thus with  $\rho_T \equiv 1$  the system  $(OC_h)$  mimics the Stokes system.

#### 4. OPTIMIZATION

In the following present the optimization algorithm. As it is common in image registration we follow the multiresolution strategy  $rl = 0 \rightarrow rl = -1 \dots \rightarrow rl = -rl_{min}$ , where at each lower resolution level,  $rl$ , we have to solve a smaller optimization problem (cf. Section 4.3). This problem is solved iteratively by an inexact SQP algorithm, where the initial guess is taken from the solution at the lower resolution level. At each stage of the iterative process we need to solve a possibly large KKT system and this is done by a Krylov subspace method with a suitable preconditioner.

**4.1. One level image registration.** For a fixed resolution level  $rl$  let  $\rho_R^{rl}, \rho_T^{rl}$  be the input mass density functions expected to be registered. The resolution of images determine uniquely the shape of discrete domains  $\Omega_c$  and  $\Omega_u$  where we set up the discrete analogue of the objective function to be minimized and the constraints (cf. Section 3). Thus we have a nonlinear optimization problem of type  $(P_h)$  and wish to find an optimal solution  $u_{rl}^h$ . For notational convenience we omit the subscript  $_{rl}$  indicating the current resolution level. In order to solve  $(P_h)$  we use the framework of Sequential Quadratic Programming (SQP) [37]. Due to the absence of inequality constraints, the SQP method may be interpreted as a kind of Newton method applied to the optimality conditions  $(OC_h)$  derived from  $(P_h)$  (cf. [6]).

Let  $u_0^h \in \mathbb{R}^n$  be an initial guess for the displacement, typically being the prolongation from the previous resolution level or set to zero for the coarsest space, and  $p_0^h \in \mathbb{R}^m$  be an initial guess for the Lagrange multiplier. The SQP method consist of iteratively updating the approximate displacement  $u_k^h$  by an appropriate correction vector  $\delta u_k^h$  in the primal space  $\mathbb{R}^n$  and  $p_k^h$  by a vector  $\delta p_k^h$  in the dual space  $\mathbb{R}^m$ . This can be obtained by solving the KKT system

$$(14) \quad \mathcal{K}_k \delta_k := \begin{pmatrix} W_k & B_k^\top \\ B_k & \end{pmatrix} \begin{pmatrix} \delta u_k^h \\ \delta p_k^h \end{pmatrix} = - \begin{pmatrix} A^h u_k^h + B_k^\top p_k^h \\ c^h(u_k^h) \end{pmatrix} =: -b_k,$$

for one level of the SQP algorithm, where  $\mathbb{R}^{m \times n} \ni B_k := \mathbf{B}^h(u_k^h)$  and  $W_k \in \mathbb{R}^{n \times n}$  is an approximation of the Lagrangian Hessian

$$W_{k,ij} \approx A_{ij}^h + \sum_{l=1}^m \frac{\partial^2 c_i^h}{\partial u_i \partial u_j} p_{k,l}^h.$$

The exact Hessian requires second-order derivatives of the constraints which are expensive to compute. Therefore, we omit this term in our algorithm, i.e., we simply set  $W_k = A^h$ .

**Remark 4.** Observe that whenever  $c_i^h$  is approximately an affine function or  $p_k^h$  is very small, this approximation is almost accurate. On the other hand, by omitting this term we lose second-order information which may affect the rate of convergence. Nevertheless, in our experience, computing the second-order derivatives of  $c^h$  is more costly than simply requiring more iterations to obtain a desired tolerance with  $W_k = A^h$ .

At the lower resolution level, the KKT system (14) may be solved by a direct method, for example, the LU factorization. Unfortunately, when  $m, n \gg 1$  direct solvers call for a large amount of memory that may be not realizable on a common computer. We therefore allow for an approximate solutions of (14) obtained by a preconditioned generalized residual method (cf. Section 4.2). Additionally, we incorporate a specific stopping criteria to compromise the computational cost and convergence [6, Algorithm: Inexact SQP with Gmres and Smart Tests].

After an approximated solution  $\delta_k = (\delta u_k^h, \delta p_k^h)$  is obtained, we use a globalization strategy given by a line-search, where the next iterate is obtained by updating the previous one with an appropriate scaling of the computed corrections, i.e.,

$$\begin{pmatrix} u_{k+1}^h \\ p_{k+1}^h \end{pmatrix} = \begin{pmatrix} u_k^h \\ p_k^h \end{pmatrix} + \tau_k \begin{pmatrix} \delta u_k^h \\ \delta p_k^h \end{pmatrix},$$

where  $\tau_k$  is determined by the backtracking Armijo algorithm satisfying the condition

$$(15) \quad g(u_k^h + \tau_k \delta u_k^h) \leq g(u_k^h) - \eta \tau_k \nabla g(u_k^h) \cdot \delta u_k^h, \quad 0 < \eta < 1,$$

where  $g$  is an appropriately chosen SQP merit function which plays an extremely important role in the optimization process. To achieve convergence for convex problems one may simply use the merit function  $g(u_k^h) = \|b_k\|^2$ , where  $b_k$  is defined in (14). Nonconvex problems, however, require more complex merit functions to guarantee convergence. Here, we choose the augmented  $l_2$  version

$$(16) \quad g(u^h, \sigma) = S^h(u^h) + \langle p^h, c^h(u^h) \rangle + (\sigma/2) \|c^h(u^h)\|_2^2,$$

where  $\sigma > 0$  is an appropriately chosen penalty parameter [37].

**Remark 5.** In order to avoid foldings we extend the Armijo algorithm by a ‘diffeomorphic test’ [34]. We have experienced that for irregular images the approximated correction vectors may cause the solution to be nondiffeomorphic, i.e., foldings may appear in the deformed grid. This happens whenever the approximated Jacobian determinant changes sign. For this reason, we additionally require

$$(17) \quad \min_{x \in \Omega_c} V^h(u^h)(x) \geq \delta,$$

for some fixed  $\delta > 0$ , where  $V^h$  is an approximation of the determinant defined in (30). More precisely, we choose  $\tau_k$  sufficiently small so as to satisfy both (15) and (17).

Finally, we provide the general scheme for solving  $(P_h)$ .

**Algorithm 1** (One level image registration).

Given:  $u_0^h, p_0^h$ , stopping criteria

Initialize: set  $k = 0$

While the stopping criteria is not satisfied:

- (1) set up the KKT system (14) for  $(u_k^h, p_k^h)$ ,
- (2) compute an approximate solution  $(\delta u_k^h, \delta p_k^h)$  of (14),
- (3) apply the backtracking Armijo line search to obtain  $\alpha_k$  satisfying (15),
- (4) update  $(u_{k+1}^h, p_{k+1}^h) = (u_k^h, p_k^h) + \alpha_k (\delta u_k^h, \delta p_k^h)$ ,
- (5) update  $k = k + 1$ ,

**4.2. KKT solver.** Our optimization method described in the previous section is based on the assumption that the direct solution of the KKT system (14) is impossible or hard to obtain. If we are dealing with high resolution images (particularly in 3D) our problem may easily reach a few million unknowns. Therefore, we apply the generalized minimal residual method (GMRES) which is both ‘relatively cheap’ and stable. As known, the GMRES is guaranteed to be convergent in at most  $N$  iterations where  $N$  is the matrix size. Naturally performing  $N$  iterations is not cheap in the sense of the required computational time. Nevertheless, we expect to obtain a reasonable approximation after a relatively small number of iterations. Furthermore, since our KKT system is ill-conditioned, we cannot expect fast convergence. Hence, there is a need for a good preconditioner. In [1] the authors tested different preconditioning methods and concluded that the best result is obtained by applying the block triangular preconditioner

$$(18) \quad \mathcal{P} = \begin{pmatrix} A & \\ B & -C \end{pmatrix}, \quad \text{where} \quad C = BA^{-1}B^\top - \text{Schur complement matrix.}$$

Note, that applying the inverse of  $\mathcal{P}$  from the left we get a preconditioned matrix

$$\mathcal{K}_{\mathcal{P}} \approx \begin{pmatrix} I & A^{-1}B^\top \\ & I \end{pmatrix}$$

which is upper triangular. Thus, one sweep of Gaussian elimination leads to the required solution. Unfortunately, obtaining the exact inverse of  $\mathcal{P}$  is almost as costly as inverting the original matrix  $\mathcal{K}$  itself. This necessitates the need for a cheap approximate inverse of  $A$  and  $C$ .

**4.2.1. Approximate inverse of  $A$ .** Since the matrix  $A$  corresponds to the elliptic Navier–Lamé operator, one can show that a multigrid solver provides an excellent convergence rate, providing the *Poisson ratio*  $\nu > 0$  is relatively small. For  $\nu \rightarrow 0.5$  the rank deficient  $\nabla^h \nabla^h \cdot$  operator becomes dominating and  $A$  becomes nearly singular [49]. In particular, the solution to the linear elastic system may be perturbed by any divergence-free displacement without affecting the residual. These perturbations may have highly oscillatory behavior and thus the produced errors cannot be smoothed efficiently by the coarser grid information. Since most of the soft human tissues exhibit near incompressible behavior, we reformulate the linear elasticity problem into an equivalent form that does not suffer from the near singularity of the original equation [20].

For a given  $u^h$  let us introduce a new variable  $q^h \in H_c$  to be  $q^h = -\lambda \nabla^h \cdot u^h$ . We now rewrite the discrete Navier–Lamé operator (33) as

$$(19) \quad A_{aug}^h = \begin{pmatrix} \mu \nabla^h \times \nabla^h \times - 2\mu \nabla^h \nabla^h \cdot & \nabla^h \\ \nabla^h \cdot & \lambda^{-1} \end{pmatrix}.$$

For later use we define a distribution matrix

$$(20) \quad M^h = \begin{pmatrix} I & -\nabla^h \\ \mu \nabla^h \cdot & -2\mu \nabla^h \cdot \nabla^h \end{pmatrix}.$$

Using the properties of our discrete operators we can easily observe that

$$(21) \quad A^h M^h = \begin{pmatrix} \mu(\nabla^h \times \nabla^h \times - \nabla^h \nabla^h \cdot) & \\ (1 + \frac{\mu}{\lambda}) \nabla^h \cdot & -(1 + \frac{2\mu}{\lambda}) \nabla^h \cdot \nabla^h \end{pmatrix}$$

which is a block triangular matrix with Laplace-like operators on the main diagonal. For such matrices a standard Gauss–Seidel relaxation method provides perfect smoothing properties [20]. Here, we used linear interpolation as a prolongation operator and restriction operators constructed with the following stencils

$$S_{u_1} = \frac{1}{8} \begin{pmatrix} 1 & & 1 \\ 2 & \nabla & 2 \\ 1 & & 1 \end{pmatrix}, \quad S_{u_2} = \frac{1}{8} \begin{pmatrix} 1 & 2 & 1 \\ & \triangleright & \\ 1 & 2 & 1 \end{pmatrix}, \quad S_p = \frac{1}{4} \begin{pmatrix} 1 & & 1 \\ & \bullet & \\ 1 & & 1 \end{pmatrix}.$$

We experienced that such a multigrid solver with a  $V(1, 1)$  cycle is mesh independent. Furthermore it does not depend on the elasticity parameters, thereby allowing for a nearly incompressible setting.



4.2.2. *Approximate inverse of  $S$ .* For the efficient solver for the Schur complement system with the matrix  $C = BA^{-1}B^\top$  we follow the strategy based on approximate commutators [4]. This approach consists of finding an approximate solution to the matrix equation

$$B^\top X = AB^\top.$$

Under the assumption that  $B$  has full row rank, the least square solution to the above system is

$$X \approx (BB^\top)^{-1}BAB^\top.$$

Then we can approximate the inverse of the Schur complement matrix by

$$(22) \quad C^{-1} \approx (BB^\top)^{-1}BAB^\top(BB^\top)^{-1}.$$

By definition the matrix  $BB^\top$  is positive semidefinite and strictly positive definite whenever  $B$  has full row rank. Unfortunately,  $BB^\top$  is typically ill-conditioned and thus iterative methods devoted to symmetric positive definite systems, converge slowly. As suggested in [1] we use the complete LU factorization. Note, that the size of this matrix is  $m \approx \frac{1}{3}N$  and is thus much cheaper than the LU factorization of the original problem. Having the *LU* form of  $BB^\top$ , the application of the inverse Schur complement becomes straight forward.

4.3. **Multilevel approach.** Due to the fact that our constraints  $c_i^h$ ,  $i = 1, \dots, m$  are nonconvex functions, the problem  $(OC_h)$  may possess many different local solutions. Although we cannot guarantee that our algorithm converges to the global minimizer, we can decrease the risk of ending up with a stationary point being far away from it. This is obtained by the multilevel strategy seen in Figure 1. This approach consist of solving a sequence of subproblems with coarser images. These

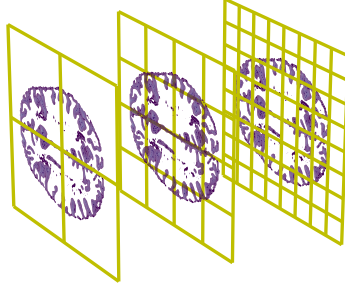


FIGURE 1. Multiresolution coarsening, from left to right

are obtained from the original images by applying a smoothing filter. Starting with the coarsest images we apply the image registration algorithm and then interpolate the obtained solution on the finer domain. One easily observes that the smoother the images are the more regular the problem is and thus providing a higher chance in finding the global solution. Moreover, this approach significantly accelerates and stabilizes the solution algorithm as it provides an excellent initial guess for the current resolution problem.

Here we apply a standard filter based on the stencil  $S_p$  described in Section 4.2.1. Let  $I_{low}$  denote the restriction operator to a lower resolution level. Then the overall image registration algorithm is expressed as follows.

**Algorithm 2.**

Given: original images  $R, T$ , integer  $rl_{min}$

Initialize:  $k = rl_{min}$ ,  $u_0^h = 0$ ,  $p_0^h = 0$

- (1) produce lower resolution images  $R_k = I_{low}^k R$ ,  $T_k = I_{low}^k T$
- (2) set up the image registration problem on the resolution level  $-k$
- (3) apply Algorithm 1 to get  $(u_k^h, p_k^h)$
- (4) prolongate  $(u_k^h, p_k^h) \rightarrow (u_{k-1}^h, p_{k-1}^h)$
- (5) update  $k = k - 1$
- (6) if  $k \geq 0$  go to step 1

## 5. NUMERICAL EXAMPLES

Next, we present several examples to demonstrate the advantages of our approach. Among the performed tests we consider a tailored example (Ex1) and a real world example (Ex2). The first case is to show how our algorithm works with an 'ideal' input data, while the second example contains perturbed data requiring some preprocessing.

In order to compare different results we provide two quantitative information, namely the value  $\text{Elas} := S^h(u^h)$  which approximates the elastic strain energy and  $\text{DMP} = \|c^h(u^h)\|_\infty$  which locally measures the distance of the determined deformed image to the reference one.

**5.1. Tailored example (Ex1).** Here we demonstrate that our algorithm recovers the true deformation. This example is taken from [19]. Let us define a real function

$$q(z) := \left( -\frac{1}{8\pi}z^2 + \frac{1}{256\pi^3} + \frac{1}{32\pi} \right) \cos(8\pi z) + \frac{1}{32\pi^2}z \sin(8\pi z).$$

Now let  $\Omega = [-0.5, 0.5] \times [-0.5, 0.5]$  and set  $\rho_T$  to be uniformly one, i.e.,  $\rho_T = 1$  on  $\Omega$ . Let us define a deformation  $\phi_e: \Omega \rightarrow \Omega$  as

$$\phi_e(x) = \begin{pmatrix} x_1 + 4q'(x_1)q(x_2) \\ x_2 + 4q(x_1)q'(x_2) \end{pmatrix}.$$

We set  $\rho_R$  to be

$$\rho_R(x) = \det(\nabla \phi_e(x))(\rho_T \circ \phi_e)(x) = \det(\nabla \phi_e(x)).$$

The above setting guarantees that the mass preserving constraint is exactly satisfied for the displacement field

$$u_e(x) = \begin{pmatrix} 4q'(x_1)q(x_2) \\ 4q(x_1)q'(x_2) \end{pmatrix}.$$

On the other hand we know that there may exist other mass preserving solutions with possibly lower elastic strain energy. Therefore we cannot expect that our algorithm produces approximate solutions that converge to  $u_e$ . This solution determines an upper bound for the value  $\text{Elas}$ , namely  $\text{Elas}_{\max} = S^h(u_e^h)$ , where  $u_e^h$  stands for the discretization of  $u_e$ , and a lower bound for the dissimilarity measure  $\text{DMP}_{\min} = 0$ . We expect to find an approximate displacement with  $\text{Elas} \leq \text{Elas}_{\max}$  and  $\text{DMP} \geq 0$ .

In Figure 2, we present the input data together with the known deformation field and the corresponding difference image. In this example we discretize  $\Omega$  over a grid of  $64 \times 64$  cells.

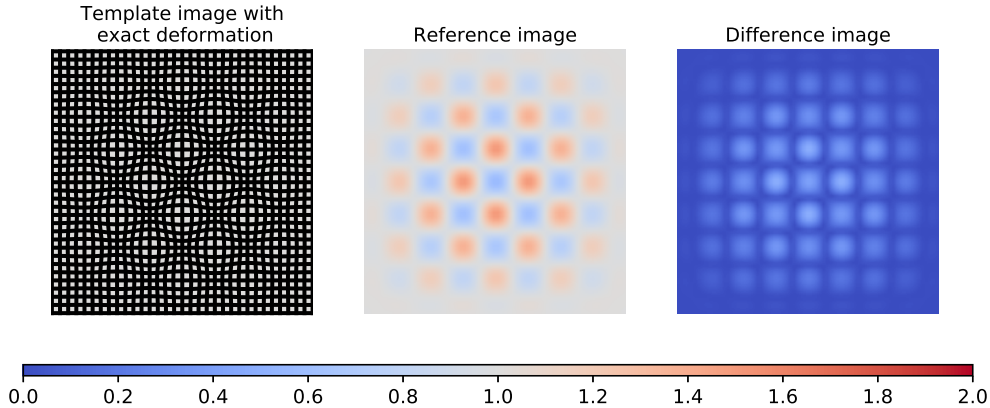


FIGURE 2. Example 1 - input data

Further, we calculate

$$\text{DE} = \|u^h - u_e^h\|_\infty,$$

where  $u^h$  stands for the computed approximation, to measure how the minimization of  $S^h$  influences the solution.

Figure 3 illustrates how the specific properties of the approximated mass preserving solution  $u_{mp}^h$  changes during the optimization process. For comparison purposes, we also depict the results obtained by solving the regularized problem (1) with the dissimilarity measure defined in (2) for different values of the regularization parameter  $\alpha$ . We call the approximate solution obtained with this method  $u_\alpha^h$ . In this example we skip the multiresolution strategy and stop the iterative solver for the KKT system (14) when the residual norm achieves the tolerance level  $r_{tol} = 1 \cdot 10^{-5}$ . The Lamé constants are set to be  $\mu = 1$ ,  $\lambda = 0$ .

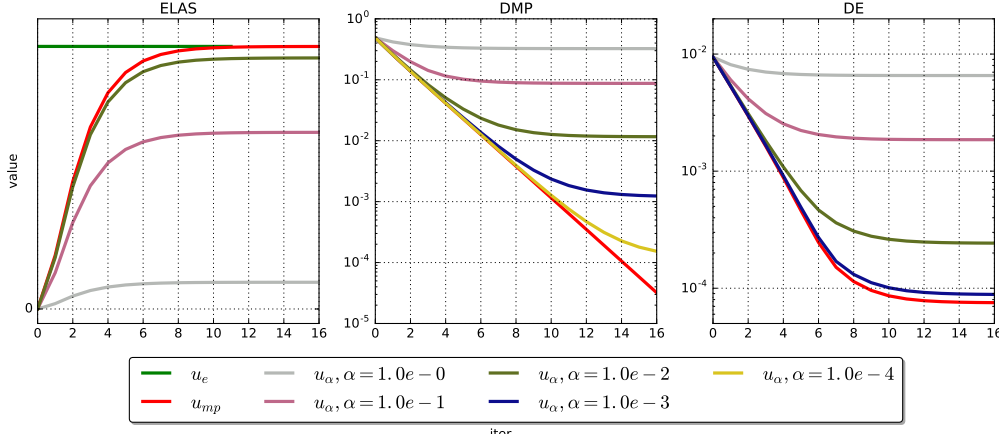


FIGURE 3. Example 1 - Elas - DMP - DE

For  $\alpha$  tending to zero, both the approximated displacements  $u_{mp}^h$  and  $u_\alpha^h$  exhibit similar properties. Indeed, we see that for  $\alpha \ll 1$ , the curves representing the quantities Elas, DMP and DE (Figure 3 left, middle and right respectively) overlap, which is to be expected.

The main difficulty in the regularized problem is that we do not know the regularization parameter  $\alpha$  a priori. One can naturally use the L-curve or continuation method [7, 27, 47] in order to find an optimal parameter, but these approaches are rather computationally expensive. In practice, one simply chooses an artificial value. Figure 4 demonstrate how strongly this choice affects the optimization process and the solution. If  $\alpha$  is chosen too large, we obtain very smooth approximations after few iterations, whereupon the process stagnates and the solution stays far away from  $u_e$ . We obtain a reasonably good approximation for  $\alpha < 1 \cdot 10^{-3}$ . In this situation, however, we must solve extremely ill-conditioned systems. Figure 4(left) presents the required number of iterations for the GMRES solver with algebraic multigrid preconditioner [3]. In Figure 4(right) we present how the regularization affects the obtained deformed image model.

**5.2. Real world example (Ex2).** In this example we process real world data obtained from [29]. They illustrate the cortical tissue of human brain at different positions in  $z$  dimension. The image resolution is  $128 \times 128$ . We first apply a Gaussian filter on these images to get rid off the present noise. Then, in order to force the mass preserving condition, we scale these data to get intensity values in range  $[\delta, 1]$  and modify in such a way to obtain the same total mass (sum of all pixel intensities). The value  $\delta = 0.03$  prevents instabilities during the optimization process. We present the input data and the corresponding difference image on Figure 5. We see that the biggest difference appears on the boundary of the cortex tissue. This situation is very realistic and typically comes from spatial misalignment.

We adapt the algorithm to work on free resolution levels, i.e., at each resolution level  $rl$  we stop the iterations as soon as

$$\text{DMP}_{rl} < C_{rl} \cdot 10^{-3},$$

where  $C_{rl}$  corresponds to the cell size at level  $rl$ , which is relative to the cell size at the finest resolution, i.e.,  $C_0 = 1$ ,  $C_1 = 4$ ,  $C_2 = 16$ . As before we set  $\mu = 1$ ,  $\lambda = 0$ .

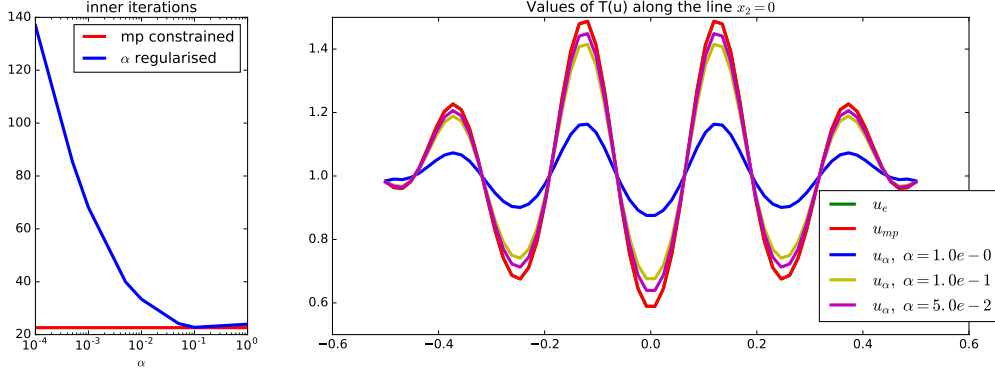


FIGURE 4. Example 1 - left: average number of inner iterations for one KKT solve, right: values of deformed image along the axis  $y = 0$

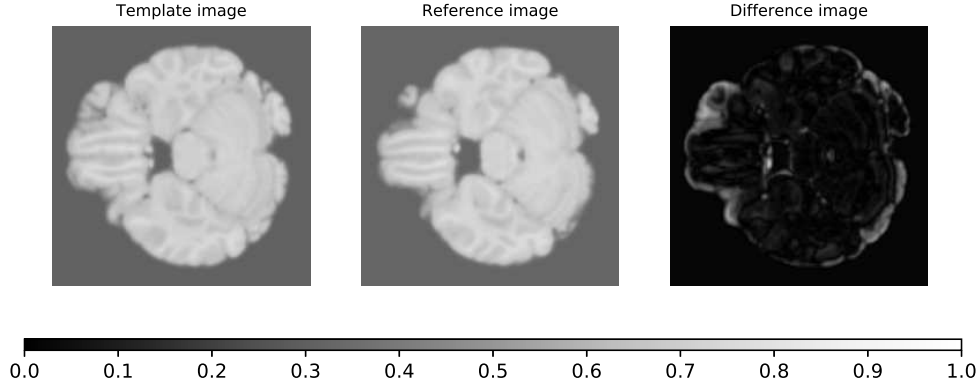


FIGURE 5. Example 2 - input data

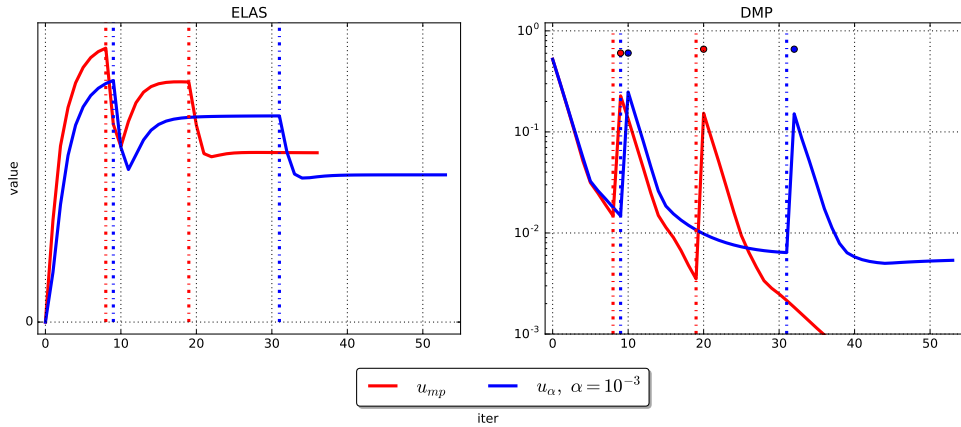


FIGURE 6. Example 2 - Elas - DMP -  $DMP_{\text{global}}$ , vertical dotted lines separate different resolution levels,  $\bullet$  denote the reference value of DMP for zero vector

We present the results on Figures 6 and 7. Information from all resolution levels are posed at one figure and separated by vertical dotted lines. Thus we can see the overall performance of the registration algorithm. We can see that DMP is reduced significantly by the solution from the

coarser grid and thus justifies the use of the multiresolution strategy. We performed 36 iterations to obtain a solution with the desired DMP tolerance. The regularized problem with parameter  $\alpha = 10^{-3}$ , however, does not lead to a solution with  $\text{DMP} < 10^{-3}$  and requires in fact a smaller  $\alpha$ . Figure 7 shows the deformation grid which allows for an almost perfect match of the images, while the mass is preserved with high accuracy.

**Remark 6.** We point out that the ratio between the Lamé constants  $\mu/\lambda$  has a big influence on the determined solution. Whenever  $\mu/\lambda$  is large the algorithm allows for solutions that are not divergence-free. On the other hand, when  $\mu/\lambda$  tends to zero, the divergence part of the elastic strain energy dominates. In this situation displacement fields with a dominating divergence-free part are more likely to be chosen.

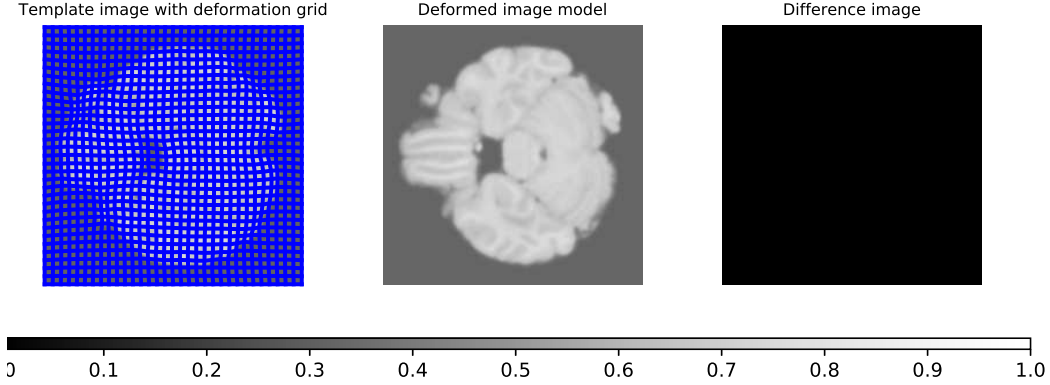


FIGURE 7. Example 2 - deformation grid - deformed image model - difference image

## 6. CONCLUSION

We considered a 2D image registration problem as an optimization problem with a fully nonlinear partial differential equation as a hard constraint. This constraint expresses the preservation of mass while the cost functional controls the linear elastic strain energy of a deformation. We proved that the problem stated in this form has a weak solution in a Sobolev space under fairly mild assumptions. Furthermore, we developed a numerical scheme to solve the problem and demonstrated its robustness on synthetic and real world data. The method involves no regularization parameter and allows to match the images accurately while preserving the mass with high precision.

## APPENDIX A. DISCRETIZATION

**A.1. Staggered grid.** We assume that the input data  $R$  and  $T$  are  $n_1 \times n_2$  matrices describing the pixels intensities. We identify each pixel with a square, whose side length is  $h$  and assume that the given data are associated with the cell centers. The unknown displacements  $u$  and the corresponding deformations  $\phi$  are discretized as follows. Let  $e_i$  be the  $i$ 'th Cartesian axis vector,  $i = 1, 2$ . We store the variables  $u_i^h$  ( $\phi_i^h$  respectively) on the center of grid edges whose orientation is consistent with  $e_i$ , where the superscript  $\cdot^h$  indicates the discrete approximation. For simplicity let us introduce the following meshes

$$\begin{aligned}
 \Omega_n &= \{(ih, jh) : i \in \{0, \dots, n_1\}, j \in \{0, \dots, n_2\}\} \\
 \Omega_c &= \{((i+1/2)h, (j+1/2)h) : i \in \{0, \dots, n_1-1\}, j \in \{0, \dots, n_2-1\}\} \\
 \Omega_{u_1} &= \{((i+1/2)h, jh) : i \in \{0, \dots, n_1-1\}, j \in \{0, \dots, n_2\}\} \\
 \Omega_{u_2} &= \{(ih, (j+1/2)h) : i \in \{0, \dots, n_1\}, j \in \{0, \dots, n_2-1\}\}.
 \end{aligned}
 \tag{23}$$

Moreover, let  $H_n, H_c, H_{u_1}, H_{u_2}$  be appropriate spaces of discrete functions defined on  $\Omega_n, \Omega_c, \Omega_{u_1}$  and  $\Omega_{u_2}$  respectively. Additionally we define  $H_u$  to be  $H_u := H_{u_1} \times H_{u_2}$ . With this notation we can write  $R, T \in H_c, \phi_i^h, u_i^h \in H_{u_i}, i = 1, 2$  and  $\phi^h, u^h \in H_u$ . It is worth to note that  $u^h$  is the optimization variable and the deformation  $\phi^h$  appears implicitly as  $\phi^h(x) = x + u^h(x), x \in \Omega_c$ .

Next, we define discrete derivatives of the quantities defined on the staggered grid  $\Omega_{u_1} \times \Omega_{u_2}$  as

$$(24) \quad \begin{aligned} D_1^i u^h(x) &= \frac{1}{h} (u_i^h(x_1 + h/2, x_2) - u_i^h(x_1 - h/2, x_2)), \quad x \in \Omega_n \text{ if } i = 1 \text{ or } x \in \Omega_c \text{ if } i = 2, \\ D_2^i u^h(x) &= \frac{1}{h} (u_i^h(x_1, x_2 + h/2) - u_i^h(x_1, x_2 - h/2)), \quad x \in \Omega_c \text{ if } i = 1 \text{ or } x \in \Omega_n \text{ if } i = 2, \end{aligned}$$

where near the boundary we utilize homogeneous Dirichlet boundary conditions. Observe that with the above second-order accurate definition, the first-order derivatives of a certain quantity in  $H_u$  are not collocated with itself. Indeed, one easily notices that  $D_i^i: H_u \rightarrow H_n$ , but  $D_i^j: H_u \rightarrow H_c$  for  $i \neq j$ . Moreover, locations of the normal and tangential derivatives differ. Such collocation of unknowns and their derivatives calls for introducing projection operators when the constraints are discretized. Therefore we define operators  $P_{u_i \rightarrow c}: H_{u_i} \rightarrow H_c, P_{u_i \rightarrow n}: H_{u_i} \rightarrow H_n, i = 1, 2$  as bilinear interpolation, e.g.

$$(25) \quad P_{u_1 \rightarrow c} u^h(x) = \frac{1}{2} (u_1^h(x_1 - h/2, x_2) + u_1^h(x_1 + h/2, x_2)), \quad x \in \Omega_c.$$

**A.2. Image model.** An important issue is the observation that the input data may have extremely irregular structure. This would lead to instabilities when a numerical method is implemented. In order to minimize the risk of failure there is a need for additional regularization. As one of the building blocks we require a continuous and differentiable representation of the image data  $T$ . The simplest and computationally cheapest choice would be the linear interpolation method. Clearly the differentiability assumption is then not satisfied. Therefore, following [25] we use the cubic B-spline interpolation scheme [46]. Due to the Curry–Schoenberg theorem [43] we can find a unique B-spline representation of any order for our discrete data  $T$ . The order 3 is chosen due to the minimum curvature property. With this choice we have the following image model

$$(26) \quad \rho_T(x) = \mathcal{I}_{bs}[T](x) := \sum_{i=0}^3 \sum_{j=0}^3 B_i(\bar{x}_1) B_j(\bar{x}_2) \gamma_{\bar{i}+i, \bar{j}+j},$$

where  $(\bar{i}, \bar{j})$  is the index related to the control cell containing  $x = (x_1, x_2)$  by the formula  $\bar{i} = \lfloor x_1/h \rfloor - 1, \bar{j} = \lfloor x_2/h \rfloor - 1$  and  $\bar{x}$  is the relative position of  $x$  inside that cell, i.e.,

$$\bar{x}_1 = x_1/h - (\bar{i} + 1), \quad \bar{x}_2 = x_2/h - (\bar{j} + 1).$$

The functions  $B_i, i = 0, \dots, 3$  are the Bernstein basis polynomials of degree 3. The coefficients  $\gamma_{i_1, i_2}, i_k = -1, \dots, n_k + 1$  may be precomputed. Observe that the continuous function  $\rho_T$  defined in this way possesses smooth derivatives up to third order which can be calculated directly in a simple and computationally efficient way. The B-splines method gives a very good compromise between the accuracy of the interpolation and the reasonable computation time.

Let  $u^h \in H_u$  be a given displacement function and  $\phi^h = id + u^h$  be its corresponding deformation. With this model at hand we can easily calculate the values of  $\rho_T \circ \phi^h$  and its derivatives at the shifted cell centers

$$(27) \quad (y_1, y_2) = (x_1 + P_{u_1 \rightarrow c} u_1^h, x_2 + P_{u_2 \rightarrow c} u_2^h), \quad (x_1, x_2) \in \Omega_c.$$

Let  $\rho_T \circ \phi^h \in H_c$  denote the warped template image uniquely determined by the displacement  $u^h$  and its derivative  $(\nabla^h \rho_T) \circ \phi^h \in H_c \times H_c$ , i.e.,

$$\begin{aligned} \rho_T \circ \phi^h(x) &= \rho_T(y), \quad x \in \Omega_c \\ (\nabla_1^h \rho_T) \circ \phi^h(x) &= \partial_{x_1} \rho_T(y), \quad x \in \Omega_c \\ (\nabla_2^h \rho_T) \circ \phi^h(x) &= \partial_{x_2} \rho_T(y), \quad x \in \Omega_c, \end{aligned}$$

where  $y = (y_1, y_2)$  is given as in (27). Then the linear operator  $\delta(\rho_T \circ \phi^h) \in \mathcal{L}(H_u, H_c)$  being the Fréchet derivative of  $\rho_T \circ \phi^h$  is

$$(28) \quad \delta(\rho_T \circ \phi^h)[\cdot] = (\nabla_1^h \rho_T \circ \phi^h) P_{u_1 \rightarrow c}[\cdot] + (\nabla_2^h \rho_T \circ \phi^h) P_{u_2 \rightarrow c}[\cdot].$$

**A.3. Constraints.** Observe that the constraint  $c$  contains the Jacobian determinant  $\det(\nabla \phi)$ , which couples all elements of the Jacobian matrix of  $\phi$ . This is in fact the most challenging element to discretize. As mentioned in Section A.1, derivatives are located at different locations. A natural solution would be to apply the projection operator. However, this approach may lead to instabilities as observed in [42]. Therefore, we follow the method proposed in [24]. Instead of discretizing  $\det(\cdot)$  in  $\Omega_c$  we calculate the approximated change of volume for each cell. This method is accurate up to second-order and allows to easily detect instabilities [42]. More precisely, let  $x \in \Omega_c$  and  $V(x)$  denote the cell centered at point  $x = (x_1, x_2)$ . Then we have

$$(29) \quad \det(\nabla \phi)(x) = \frac{1}{h^2} \text{vol}(\phi(V(x))) + \mathcal{O}(h^2) = \frac{1}{h^2} \int_{\phi(V(x))} dx + \mathcal{O}(h^2)$$

and the latter is approximated by a procedure precisely described in [24].

Now we define the mapping  $\det^h: H_u \rightarrow H_c$  by

$$(30) \quad V^h(u^h)(x) := \frac{1}{h^2} \text{vol}(\phi^h(V(x))), \quad x \in \Omega_c,$$

with  $\phi^h = id + u^h$ . From the above discussion, this yields a consistent second-order approximation of the Jacobian determinant [42]. Correspondingly, let  $\delta V^h(u^h) \in \mathcal{L}(H_u, H_c)$  denote the Fréchet derivative of  $V^h$  at  $u^h$ . Similarly, we consider the nonlinear constraint  $c^h: H_u \rightarrow H_c$  where

$$(31) \quad c^h(u^h)(x) = V^h(u^h)(x)(\rho_T \circ \phi^h)(x) - \rho_R(x), \quad x \in \Omega_c.$$

Furthermore, its Frechét derivative at  $u^h$  denoted by  $B^h(u^h) \in \mathcal{L}(H_u, H_c)$  is given by

$$(32) \quad \delta c^h(u^h)[\cdot](x) = \delta V^h(u^h)[\cdot](x)(\rho_T \circ \phi^h)(x) + \delta(\rho_T \circ \phi^h)[\cdot](x) V^h(u^h)(x), \quad x \in \Omega_c.$$

**A.4. Objective functional.** Note, that the Fréchet derivative of  $\mathcal{S}$  in  $H^1(\Omega; \mathbb{R}^2)$  is

$$d\mathcal{S}(u)[v] = \langle \mu \Delta u + (\mu + \lambda) \nabla(\nabla \cdot u), v \rangle_{H^{-1}, H^1} =: \langle \mathcal{A}(u), v \rangle_{H^{-1}, H^1} \quad \text{for all } v \in H^1(\Omega; \mathbb{R}^2),$$

where  $\Delta$  is the Laplace operator and  $\nabla \cdot$  denotes divergence operator. A straightforward approach would be to discretize the operator  $\mathcal{A}$  by  $A^h$  and define  $S^h(u^h)$  to be  $S^h(u^h) := \frac{1}{2} \langle u^h, A^h u^h \rangle$  as in [28]. This approach, however, has a significant drawback which causes the multigrid method to be inefficient for  $\lambda \gg \mu$ . Therefore, we choose a mimetic discretization (cf. [20]) in order to preserve the properties of the continuous operator. Let  $\nabla \times$  denote the self-adjoint curl operator. Then for any continuous vector field  $u$  and scalar field  $f$  we have the identities:

- (i)  $\Delta u = \nabla(\nabla \cdot u) - \nabla \times (\nabla \times u)$
- (ii)  $\nabla \times (\nabla f) = 0$
- (iii)  $\nabla \cdot (\nabla \times u) = 0$

Using the property (i) we can write

$$d\mathcal{S}(u)[v] = \langle \mu \nabla \times (\nabla \times u) - (\lambda + 2\mu) \nabla(\nabla \cdot u), v \rangle_{H^{-1}, H^1} =: \langle \mathcal{G}^* \mathcal{G} u, v \rangle_{H^{-1}, H^1},$$

where  $\cdot^*$  denotes the adjoint operator and

$$\mathcal{G} = \begin{pmatrix} \sqrt{\mu} & 0 \\ 0 & \sqrt{2\mu + \lambda} \end{pmatrix} \begin{pmatrix} \nabla \times \\ \nabla \cdot \end{pmatrix}.$$

Following [20] we assemble operators  $\nabla^h \times$  and  $\nabla^h \cdot$  satisfying properties (ii) and (iii) on the discrete spaces, and set

$$(33) \quad G^h := \begin{pmatrix} \sqrt{\mu} & 0 \\ 0 & \sqrt{2\mu + \lambda} \end{pmatrix} \begin{pmatrix} \nabla^h \times \\ \nabla^h \cdot \end{pmatrix}, \quad A^h := h^2 G^{h, \top} G^h.$$

Finally, our mimetic discretization of the objective functional reads

$$(34) \quad S^h(u^h) = \frac{1}{2} \langle u^h, A^h u^h \rangle,$$

whose Fréchet derivative is given by

$$(35) \quad \delta S^h(u^h)[v^h] = \langle u^h, A^h v^h \rangle \quad \forall v^h \in H_u.$$

## REFERENCES

- [1] O. Axelsson and J. Karatson. Efficient preconditioned solution method for elliptic partial differential equation. *Bentham Science*, 2011.
- [2] R. Bajcsy and S. Kovačič. Multiresolution elastic matching. *Computer Vision, Graphics and Image Processing*, 46:1–21, 1989.
- [3] W. N. Bell, L. N. Olson, and J. Schroder. Pyamg: Algebraic multigrid solvers in python. <http://www.pyamg.org>, 2013. Version 2.1.
- [4] M. Benzi, G.H. Golub, and J. Liesen. Numerical solutions of saddle point problems. *Acta Numerica*, 14:1–137, May 2005.
- [5] M. Burger, J. Modersitzki, and L. Ruthotto. A hyperelastic regularization energy for image registration. *SIAM Journal on Scientific Computing*, 35(1):B132–B148, 2013.
- [6] R. Byrd, F. Curtis, and J. Nocedal. An inexact sqp method for equality constrained optimization. *SIAM Journal of Optimization*, 19(1):351–369, 2008.
- [7] D. Calvetti, S. Morigi, L. Reichel, and F. Sgallari. Tikhonov regularization and l-curve for large discrete ill-posed problems. *Journal of Computational and Applied Mathematics*, 123(1–2):423–446, 2000.
- [8] G.E. Christensen, R. Rabbit, and M.I. Miller. Deformable templates using large deformations kinematics. *IEEE Transactions on Medical Images*, 5(10), 1996.
- [9] N. Chumchob and K. Chen. A robust affine image registration method. *International Journal of Numerical Analysis and Modeling*, 6(2):311–334, 2009.
- [10] P.G. Ciarlet. *Mathematical Elasticity*, chapter 1: Three Dimensional elasticity. Elsevier Science Publishers B. V., 1988.
- [11] B. Dacorogna. *Direct methods in the calculus of variations*. Applied Mathematical Sciences. Springer Berlin Heidelberg, New York, 2 edition, 2008.
- [12] B. Dacorogna and J. Moser. On partial differential equations involving the jacobian determinant. *Annales de l’Henri Poincaré Analyse non linéaire*, 7(1):1–26, 1990.
- [13] A.W. Date. *Introduction to Computational Fluid Dynamics*. Cambridge University Press, 2005.
- [14] L. Desvillettes and C. Villani. On a variant of korn’s inequality arising in statistical mechanics. *ESAIM: Control, Optimization and Calculus of Variations*, 8:603–619, 2002.
- [15] L. Evans. Partial differential equations and monge-kantorovich mass transfer - lecture notes. Berkeley, 2001.
- [16] B. Fisher and J. Modersitzki. Fast diffusion registration. *Contemporary Mathematics*, 313:117–129, 2002.
- [17] B. Fisher and J. Modersitzki. Curvature based image registration. *Journal of Mathematical Imaging and Vision*, 18(1):81–85, 2003.
- [18] B. Fisher and J. Modersitzki. A unified approach to fast image registration and a new curvature based registration technique. *Linear Algebra and its Applications*, 23(7):107–124, 2004.
- [19] B.D. Froese. A numerical method for the elliptic monge-ampère equation with transport boundary conditions. *SIAM Journal of Scientific Computing*, 34(3):1432–1459, 2012.
- [20] F. Gaspar, J. Gracia, F. Lisbona, and C. Oosterlee. Distributive smoothers in multigrid for problems with dominating grad-div operators. *Numerical Linear Algebra with Applications*, 15:661–683, 2008.
- [21] F. Gigengack, L. Ruthotto, M. Burger, C.H. Wolters, X. Jiang, and K.P. Schafers. Motion correction in dual gated cardiac pet using mass-preserving image registration. *IEEE transactions on medical imaging*, 31(3):698–712, 2012.
- [22] M.D. Gunzburger. *Perspectives in flow controll and optimization*. SIAM, Philadelphia, 2003.
- [23] E. Haber, R. Horesh, and J. Modersitzki. Numerical optimization for constrained image registration. *Numerical Linear Algebra with Applications*, 17:343–359, 2010.
- [24] E. Haber and J. Modersitzki. Numerical methods for volume preserving image registration. *Inverse Problems*, 20(5), 2004.
- [25] E. Haber and J. Modersitzki. A multilevel method for image registration. *SIAM Journal of Scientific Computing*, 27(5):1594–1607, 2006.
- [26] J.V. Hajnal, D.L.G. Hill, D.J. Hawkes, and editors. *Medical Image Registration*. CRC Press, Boca Raton, 2001.
- [27] P.C. Hansen. *Rank-Deficient and Ill-Posed Problems: Numerical Aspects of Linear Inversion*. SIAM, Philadelphia, 1997.
- [28] S. Henn and K. Witsch. Iterative multigrid regularization techniques for image matching. *SIAM Journal of Scientific Computing*, 23(4):1077–1093, 2001.
- [29] kkk. Brainweb: Simulated brain database. <http://brainweb.bic.mni.mcgill.ca/brainweb>. Accessed: 2016-04-14.
- [30] J. Kybic and M. Unser. Fast parametric elastic image registration. *IEEE RANSCTIONS ON IMAGE PROCESSING*, 12(11), 2003.



- [31] L. Lemieux, U.C. Wieshmann, N.F. Moran, D.R. Fish, and S.D. Shorvon. The detection and significance of subtle changes in mixed-signal brain lesions by serial mri scan matching and spatial normalization. *Medical Image Analysis*, 2:227–242, 1998.
- [32] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Images*, 16(2), April 1997.
- [33] V.R.S. Mani and S. Arivazhagan. Survey of medical image registration. *Journal of Biomedical Engineering and Technology*, 1(2):8–25, 2013.
- [34] J. Modersitzki. *Numerical Methods for Image Registration*. Oxford University Press, New York, 2004.
- [35] J. Modersitzki. Flirt with rigidity - image registration with a local non-rigidity penalty. *International Journal of Computer Vision*, 76:153–163, 2008.
- [36] Andriy Myronenko, Xubo Song, and Miguel A. Carreira-Perpinan. Non-parametric Image Registration Using Generalized Elastic Nets. In Xavier Pennec and Sarang Joshi, editors, *1st MICCAI Workshop on Mathematical Foundations of Computational Anatomy: Geometrical, Statistical and Registration Methods for Modeling Biological Shape Variability*, pages 156–163, Copenhagen, Denmark, 2006.
- [37] J. Nocedal and S.J. Wright. *Numerical Optimization*. Springer-Verlag, New York, 1999.
- [38] C. Pöschl, J. Modersitzki, and O. Scherzer. A variational setting for volume constrained image registration. *Inverse Problems and Imaging*, 4(3), 2010.
- [39] T. Rehman. *Efficient Numerical Method for Solution of  $L^2$  Optimal Mass Transprt Problem*. PhD thesis, School of Electrical and Computer Engineering, 2010.
- [40] K. Rohr, H. S. Stiehl, R. Sprengel, W. Beil, T. M. Buzug, J. Weese, and M. H. Kuhn. *Point-based elastic registration of medical image data using approximating thin-plate splines*, pages 297–306. Springer Berlin Heidelberg, Berlin, Heidelberg, 1996.
- [41] L. Ruthotto. Mass-preserving registration of medical images. Master’s thesis, Fachbereich Mathematik und Informatik, Westfälische Wilhelms-Universität Münster, Deutschland, 2010.
- [42] L. Ruthotto. *Hyperelastic Image Registration - Theory, Numerical Methods and Applications*. PhD thesis, Fachbereich Mathematik und Informatik, Westfälische Wilhelms-Universität Münster, Deutschland, 2012.
- [43] I.J. Schoenberg and A. Sharma. Cardinal interpolation and spline fucntions v. the b-splines for cardinal hermite interpolation. *Linear Algebra and its Applications*, 7(1):1–42, 1973.
- [44] S. Suhr. *Variational Methods for Combined Image and Motion Estimation*. PhD thesis, Fachbereich Mathematik und Informatik, Westfälische Wilhelms-Universität Münster, Deutschland, 2015.
- [45] J.P. Thirion. Non-rigid matching using demons. In *Computer Vision and Pattern Recognition, 1996. Proceedings CVPR ’96, 1996 IEEE Computer Society Conference on*, pages 245–251, 1996.
- [46] M. Unser. Splines: A perfect fit for signal and image processing. *IEEE Signal Processing Magazine*, 16(6):22–38, 1999.
- [47] C.R. Vogel. *Computational Methods for Inverse Problem*. SIAM, Philadelphia, 2002.
- [48] L. Zhu, Y. Yang, S. Haker, and A. Tannenbaum. An image morphing technique based on optimal mass preserving mapping. *IEEE Transactions on Image Processing*, 16(6), 2007.
- [49] Y. Zhu, E. Sifakis, J. Teran, and A. Brandt. An efficient multigrid method for the simulation of high-resolution elastic solids. *ACM Transactions on Graphics*, 29(16), 2010.

(Jarosław Wlazło)

TRANSPORT PROCESSES, FRAUNHOFER ITWM, FRAUNHOFER-PLATZ 1 KAISERSLAUTERN, 67663, GERMANY

*E-mail address:* `jaroslaw.wlazlo@itwm.fraunhofer.de`

(Robert Feßler)

TRANSPORT PROCESSES, FRAUNHOFER ITWM, FRAUNHOFER-PLATZ 1 KAISERSLAUTERN, 67663, GERMANY

*E-mail address:* `robert.fessler@itwm.fraunhofer.de`

(René Pinnau)

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN, DEPARTMENT OF MATHEMATICS, ERWIN-SCHRÖDINGER-STRASSE, 67663 KAISERSLAUTERN, GERMANY

*E-mail address:* `pinnau@mathematik.uni-kl.de`

(Norbert Siedow)

TRANSPORT PROCESSES, FRAUNHOFER ITWM, FRAUNHOFER-PLATZ 1 KAISERSLAUTERN, 67663, GERMANY

*E-mail address:* `norbert.siedow@itwm.fraunhofer.de`

(Oliver Tse)

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN, DEPARTMENT OF MATHEMATICS, ERWIN-SCHRÖDINGER-STRASSE, 67663 KAISERSLAUTERN, GERMANY

*E-mail address:* `tse@mathematik.uni-kl.de`